

**AD-A228 948**

**DTIC FILE COPY**

AD \_\_\_\_\_

**CONTRACT NO: DAMD17-87-C-7187**

**TITLE: MOLECULAR BIOLOGY OF STLV-III AND HTLV-IV**

**PRINCIPAL INVESTIGATOR: Joseph Sodroski, Ph.D.**

**CONTRACTING ORGANIZATION: Laboratory of Biochemistry  
Pharmacology  
Dana-Farber Cancer Institute  
44 Binney Street  
Boston, MA 02115**

**REPORT DATE: August 10, 1989**

**TYPE OF REPORT: Annual Report**

**DTIC  
ELECTE  
NOV 20 1990  
S B D**

**PREPARED FOR: U.S. ARMY MEDICAL RESEARCH AND DEVELOPMENT COMMAND  
Fort Detrick, Frederick, Maryland 21702-5012**

**DISTRIBUTION STATEMENT: Approved for public release;  
distribution unlimited**

**The findings in this report are not to be construed as an  
official Department of the Army position unless so designated by  
other authorized documents.**

**90 11 19 224**

## REPORT DOCUMENTATION PAGE

Form Approved  
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION <b>Unclassified</b>			1b. RESTRICTIVE MARKINGS		
2a. SECURITY CLASSIFICATION AUTHORITY			3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution unlimited		
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE					
4. PERFORMING ORGANIZATION REPORT NUMBER(S)			5. MONITORING ORGANIZATION REPORT NUMBER(S)		
6a. NAME OF PERFORMING ORGANIZATION Dana-Farber Cancer Institute Laboratory of Biochemistry Pharmacology		6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION		
6c. ADDRESS (City, State, and ZIP Code) 44 Binney Street Boston, MA 02115			7b. ADDRESS (City, State, and ZIP Code)		
8a. NAME OF FUNDING / SPONSORING ORGANIZATION U.S. Army Medical Research & Development Command		8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER Contract No. DAMD17-87-C-7187		
8c. ADDRESS (City, State, and ZIP Code) Fort Detrick Frederick, Maryland 21702-5012			10. SOURCE OF FUNDING NUMBERS		
			PROGRAM ELEMENT NO. 63105A	PROJECT NO. 3M2 63105H29	TASK NO. AC
			WORK UNIT ACCESSION NO. 072		
11. TITLE (Include Security Classification) Molecular Biology of STLV-III and HTLV-IV					
12. PERSONAL AUTHOR(S) Joseph Sodroski, Ph.D.					
13a. TYPE OF REPORT Annual Report		13b. TIME COVERED FROM 7/1/88 TO 6/30/89		14. DATE OF REPORT (Year, Month, Day) 1989 August 10	
15. PAGE COUNT 17					
16. SUPPLEMENTARY NOTATION					
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block-number)		
FIELD	GROUP	SUB-GROUP			
06	03		Virology; Molecular Biology; Biochemistry; Human		
06	13		Immunodeficiency Virus; HTLV-IV; STLV-III; Simian AIDS; RAI		
19. ABSTRACT (Continue on reverse if necessary and identify by block number)					
<p>Molecular characterization of the SIV<sub>mac</sub> 251 clone that exhibits <u>in vitro</u> replication competence, but poor <u>in vivo</u> replication ability was initiated by complete nucleotide sequence determination. Defects in the <u>vpr</u>, <u>nef</u> and <u>env</u> genes were noted and will be the focus of future efforts to restore <u>in vivo</u> replication competence of this cloned virus.</p>					
20. DISTRIBUTION / AVAILABILITY OF ABSTRACT <input type="checkbox"/> UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS			21. ABSTRACT SECURITY CLASSIFICATION Unclassified		
22a. NAME OF RESPONSIBLE INDIVIDUAL Mrs. Virginia M. Miller			22b. TELEPHONE (Include Area Code) (301) 663-7325		22c. OFFICE SYMBOL SGRD-RMI-S

## FOREWORD

Citations of commercial organizations and trade names in this report do not constitute an official Department of the Army endorsement or approval of the products or services of these organizations.



Accession For	
NTIS GRA&I	<input checked="checked" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

Addition information for Methods section.

The pFLB10 proviral DNA was sequenced according to the following method. Oligonucleotide primers were prepared based on consensus sequences for available SIV isolates provided in the Los Alamos data bank. Approximately 85% of these primers were found to yield readable sequence information from the pFLB10 provirus. An additional set of primers was prepared based on the sequence of cDNA clones derived from SIVmac 251 - infected cells. The latter primers were found to prime readable sequencing reactions using the pFLB10 provirus in nearly 100% of the cases. In total, sequencing reactions with approximately 45 different primers were used to complete the sequence of the entire pFLB10 provirus.

Sequencing reactions were carried out according to standard protocols based upon the method of Sanger et al. Briefly, template DNA was prepared by treatment of the pFLB10 plasmid with alkali to denature the DNA, followed by strand separation on denaturing acrylamide gels. The primer was annealed to the single-stranded template DNA by heating to 65°C then slowly cooling the reaction to 35°C. The molar ratio of primer:template for most reactions ranged from 1:1 to 3:1. The annealed primer:template was mixed with the 6mM DTT, labeling mix and sequenase enzyme (Sequenase Kit, USB Corporation) and [ $\gamma$ -<sup>35</sup>S] dATP, then incubated at room temperature for 2-5 minutes. The reactions were terminated by addition of ddATP, ddGTP, ddTTP and ddCTP, incubation at 37°C for 3-5 minutes, and addition of stop solution (Sequenase Kit, USB Corporation). Samples were loaded on simultaneous 4% and 8% acrylamide sequencing gels, or in some cases, on 6% gradient gels. Each sequencing reaction typically yielded from 400 to 550 base pairs of information. Each segment of the genome was sequenced using at least two sets of non-overlapping primers.

## Molecular Analysis of an Infectious SIV<sub>mac</sub> 251 Proviral Clone

### Abstract

Molecular characterization of the SIV<sub>mac</sub> 251 proviral clone pFLB-10 was carried out. This SIV<sub>mac</sub> 251 isolate exhibits in vitro replication competence, but replicates poorly in vivo and, as a consequence is not pathogenic. Defects in the vpr, nef and env genes were noted that may be responsible for attenuations in in vivo replicative ability.

### Introduction

The pFLB-10 provirus was derived from a bacteriophage library made from human T cells infected with SIV<sub>mac</sub> 251. The pFLB-10 provirus generates infectious SIV upon DEAE-dextran transfection of HUT78 and CEMX174 lymphocytes. Virus stocks of greater than  $5 \times 10^5$  reverse transcriptase units per milliliter were used for inoculation of rhesus macaques and cynomolgus monkeys. None of the animals inoculated with the FLB-10 virus has to date demonstrated a persistent viremia or induction of disease, whereas control animals inoculated the uncloned SIV<sub>mac</sub> 251 virus have developed AIDS-like illness in the same time period. The studies in animals are of one year duration and are ongoing. We conclude that changes in the molecularly cloned FLB-10 virus compromise its ability to allow efficient replication in vivo. Here we report the progress made in molecular analysis of this deficiency in in vivo replication/pathogenesis.

### Methods

To understand the molecular characteristics of the FLB-10 virus that might be relevant to the observed decrease in in vivo replication, the infectious provirus was completely sequenced using the Sanger dideoxy technique (1). Using chemically synthesized oligonucleotides as primers and increasing the readable sequence information by using gradient urea gels and <sup>35</sup>S label, the nucleotide sequence could be obtained using a minimal number of individual priming reactions.

### Results and Discussion

The sequence of the 3' half of the FLB-10 provirus is shown in appendix 1 and is summarized in figure 1. The nucleotide sequence of the FLB-10 clone is generally similar to that of other SIV<sub>mac</sub> 251 isolates and is more similar to those isolates than to SIV's derived from mangabeys or African green monkeys. The major open reading frames in the FLB-10 provirus are as follows:

1. vpx

The vpx protein is not necessary for SIV replication, but appears to stimulate virus replication through an unknown mechanism. The vpx open reading frame is present and a potential initiator methionine is evident at the 5' end of the reading frame. The FLB-10 vpx open reading frame could encode a protein of 113 amino acids, with a proline-rich segment near the carboxyl terminus. The vpx of FLB-10 differs from the consensus SIV sequence at one residue (62) M  $\rightarrow$  K.

2. vpr

The vpr protein is not necessary for HIV-1 gene expression, but can stimulate HIV-1 replication through its effect as a promiscuous trans-activator of gene expression. Most SIV<sub>mac</sub> isolates have an open reading frame for vpr, with heterogeneity at the 3' end. The BK28 isolate encodes a 98 amino acid vpr product, while MM142 encodes a 102 amino acid vpr protein. The vpr open reading frame of the FLB-10 provirus has obviously lost the potential to encode a protein, since the methionine likely to be utilized for initiation, based on sequence similarity to other SIV<sub>mac</sub> provirus clones, is followed within eleven codons by a stop codon. Following the stop codon, the vpr open reading frame continues with strong sequence similarity to the vpr sequence of other SIV isolates, so it is likely that correction of the single stop codon will result in a full-length vpr product. This potential vpr product would be 98 amino acids long, similar in size to functional vpr proteins observed in the HIV-1 system. In addition, there are two vpr changes in FLB-10 differing from the consensus SIV sequence (47) I  $\rightarrow$  M and (77) C  $\rightarrow$  S.

3. tat and rev

The FLB-10 tat and rev open reading frames are intact, as would be expected for an infectious proviral clone. Tat is a positive trans-activator of viral gene expression, while rev is a post-transcriptional regulator of structural protein expression. The tat gene of FLB-10 differs from consensus by only two changes (27) A  $\rightarrow$  R and (75) S  $\rightarrow$  C, while the FLB-10 rev gene demonstrates no changes from the SIV consensus sequence.

4. env

The FLB-10 envelope glycoproteins are similar to other SIV<sub>mac</sub> isolates in that a premature stop codon exists in the transmembrane glycoprotein. It appears that either one of a pair of CAG residues is converted to a TAG amber codon in various SIV<sub>mac</sub> isolates passaged in human cell lines. The FLB-10 clone has a TAG CAG sequence, similar to that seen in the SIV<sub>mac</sub>251 but unlike the other SIV<sub>mac</sub> isolates. In addition, the FLB-10 has a second premature stop codon that would result in a deletion of 18 carboxy-terminal amino acids.

## 5. nef

Although premature stop codons in HIV-1 nef genes are common, most SIV molecular clones do not have obvious premature nef stop codons. Heterogeneity occurs in the carboxyl terminus of nef in different SIV isolates (from 248-299 amino acid residues). The FLB-10 nef gene, which otherwise could encode a 263 residue protein, contains at least two defects that render it non-functional. First, the nef initiator methionine codon (ATG) has been mutated to an ATA, encoding isoleucine. The next residue, a glycine, which is important for myristillic acid addition, has been preserved in the FLB-10 provirus. Second, a premature stop codon results in only a 94 amino acid product, even if the initiator methionine were present. This result explains the observation that deletions of the FLB-10 provirus in the nef region did not result in phenotypic differences in viral replication rate or in cytopathic effect.

In summary, the molecular changes that might in part account for the attenuated in vivo replication rate of an SIV<sub>mac</sub> 251 infectious virus include changes in the vpr and nef genes as well as truncation of the transmembrane envelope glycoprotein.

Plans for the next year include mutagenesis of the FLB-10 provirus to examine the effects on in vitro and in vivo replication of restoring some of the mutated vpr, env, and nef sequences. The reading frames that are intact, like vpx and vif, will be deleted to examine the resulting phenotype.

## References

1. Sanger F., S. Nicklen and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. PNAS 74: 5463.

Figure 1. Schematic diagram of the FLB-10 3' half of the provirus. The vp~~x~~, vp~~r~~, tat, rev, env and ne~~f~~ genes of the FLB-10 SIV<sub>mac</sub> 251 isolate are shown. The X's represent changes in the FLB-10 sequence that render the open reading frame products either unable to be synthesized or prematurely terminated. The dark boxes represent the two coding exons of the functional tat and rev genes of FLB-10.



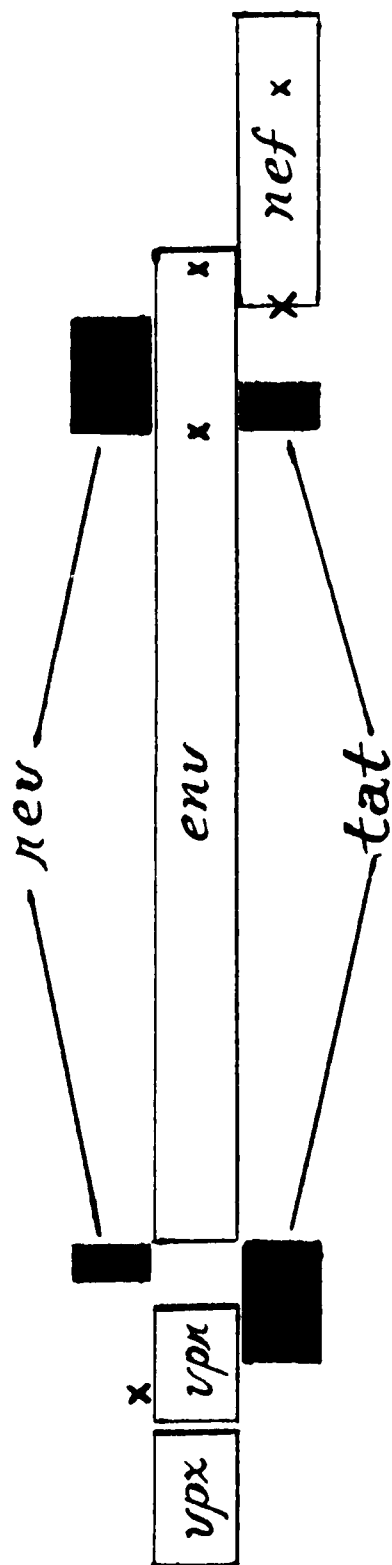


Figure 1.

Appendix 1. Complete nucleotide sequence of the 3' half of the FLB-10 provirus, with open reading frames underlined.

[illegible]

61 21 101  
TTCGGGAGAGCTCATAGGTACCCAGGTACCAAGCCTACAGTACITTAACACTAAAAGGACTT  
1: F P F A H Y Q U P S L Q Y L T L K U U  
2: G P F E L I G T R Y Q A Y S T @ H @ K @ @  
3: G P F E S @ U P R G T K P T U L N T K S S

121 141 161  
AGCGATGTCAGATCCACGGAGAGAAATCCACCTGAAACATGAGGAGAGAGACAAATAG  
S D U R S G G E N P T W N Q W R R D N R  
vpx M S D P R E R I P L G N G G F E T I  
G R C Q I P R E R S H L G T V G R R Q I

161 201 201  
 ASAGGGCCITGAAATGGCTAANCAGGAGAGTAGAGGAGATAGACAGAGAGAGGAGGATAGAG  
 1: R G L R M A K Q N S R G D K Q R G G K P  
 2: E A E E M L N R T V E I N T R E A G  
 3: E R P S N G T E G Q R E E G T R E P P G

061  
001-11A089AGCTTATTTCCAGGTTTGGAAGAAGGTCTTTGGCAATATCCTGGCATGATTA  
1: F T K G A I E P G L A K V L G I L A E D  
2: L L E E L I E Q V W Q R S L U E N Y W F D T  
3: L Y E G S L I E C R E G K S L E N Y T G M D T

100 200 300  
ACAAAGTGAATGTCACAAAGCTGTTGTGAAATACAGATACITGTTTTCATGCAAAAGTGT  
1: T T F E U T C C R I O I L U F N A K G  
2: C T G G T E S Y U N Y E Y I O L M D R A  
3: A K L S T E A M Q N T D U D C R R

361                      351                      401  
TCTTA TATTTGAGGAAATCTGTGTAGATGTCTAGGGTAABBAACCGGGGCCAGGGGCATC  
1: I Y A C G E F L Q M S R G R T R G R G M  
2: F F C F K G D R C L G E G H G A G G L  
3: L C A A K A U R V G S R D T G Q C L

421                      441                      461  
GAGAGCCAGAACCTCCTCTCTCTCCACGGACTATCATTAATGAGAGGAAGAGAGCTT  
P P P S S G S P S R T end x A M (consensus)  
G G E L P P P G L A  
S S L L P L Q T L K W R R

181 601 521  
15. AAATCAAGGAC TAAAGGGAACCATGGGGAAGTGGGCTAGTGGGAGTTTGTGGG-  
K K G T M G P M S E G B S C K  
U U U U U U U U U U U U U U U U U U

701 581 581  
 ATTGAAAGAAAGAGTTTAAAGATTTTGAATCTCTGTTGCTAAGAGCACTTCTTATCA  
 1: I E R R S F K T F C S C L A N C T W G A  
 2: L K E E A L K H F D P R L L H T A L S A  
 3: N P K K K L G N I L I L A C G L H L L V T

1: Y L V H G D T L E S A G E L I R L Q vpr  
 2: M Y N P H G D T L E S A G E L I R L Q tat  
 3: I C I I (M) E T P L E G E N S L E E S

tat initiation codon

661 681 701  
 ACGAGCGCTCTTCATGCAATTTTACGGGGGATTAACCACTCCAGAATCGGCCCACTGG  
 1: T S A I H A F E R I E P L O N R P W  
 2: R A L E M H E R G G S N H S R I G G F G vpr  
 3: N E R S S C I L E A D R T T P E S A N I tat

721 741 761  
 GGGAGGAAATCCTCTCTCACTATATCGGCGCTCTTACGGCTGCTATAACACATGCTATT  
 1: G R K S S L N Y T A L L R R A I T H A I  
 2: G G N P L S T I P P G G V L O H M L L vpr  
 3: G E E I L S Q L Y R P L E A C Y N T C Y tat

781 801 821  
 GTAAAAAGTGTCTTACCAATTCGCACTTTTCTTTTCTTAAAAAAGGGAGCTGGGGCTCTCT  
 1: V K S V A T I A S F V F L K R D W F W  
 2: Q K V L L F L P V L E S E K G T G M L  
 3: C K K C D Y H C Q F C F L K K G L G I tat

Initiation  
 codon rev

841 861 881  
 ATGAGCAGTCTCTGAAAGAGAGAAAGAACTCCGAAAGAGGCTAAGGCTAATATATCTCT  
 1: (M) C S H E F E E L R K R L I W L I env  
 2: P A U T K E K N S E K G Q G G I C  
 3: Y E G S R K S R F T P K K A N A N T S tat

901 921 941  
 CATCAAAACAACTAAGTATGGGATGCTCTTGGGAATCAGCTGCTTATGCGATCTCT  
 1: H Q T E K Y G M S W E S A A Y F F L F R rev  
 2: I K P U S (M) G C L G N G L L I A I L L C env  
 3: A S S G V W D V L G I S C L S P E C tat

981 991 1001  
 AAGTGTCTATGCACTCTATTGTACTCAATATGTTACAGTCTTTTATGCTCTATGCTG  
 1: K C L W P L L Y S I C L E L L W C D R L  
 2: S V Y G I Y C T D Y U T V E Y G V P A W env  
 3: Q V S K G S I V L N M S G S F A U Y C

1021 1041 1061  
 GAGGAATGGGACAATTCCTCTCTCTGTCAGCAAGAGAGTATGGGATACCTTGGGAATAAC  
 1: E E C D N S P L L C N Q E G G Y L G A N  
 2: R N A T I P L E C A T K S R R T W S T env  
 3: G G M R D F P S S V G P R V G I L G E A

1081 1101 1121  
 TCGGTGCGTACCAAGATA TGGTGATTAATTAGATCTGCGCCCTTAAATGTTACAGAACT  
 1: S V P T F G W P L T R I G P G C Y R A L  
 2: Q C L P F W G D Y S E I A I V V F E A env  
 3: L S A Y Q M V I I G N W P L M L G A

1141 1161 1181  
 TGATGCTTTTCAAGATACAGTACAGAAAGAGCTATAGAGGAGGATATGCAAACTCTTGA  
 1: P C L H I C L P L P L P L P L P L P L P L  
 2: P A L E N T V T F G A I E D V W Q L F E env  
 3: L M L P F I D S C N D C P T Y G R C

1201 1221 1241  
 GAECTCAATAAAGCCTTGTGTAAAAATTATCCCCATTATGCATTACTATGAGATCCAAATA  
 1: D L N K A L C K I P I M H Y Y E M Q Q  
 2: T S I K P C V K L S P L C I T M R C N K env  
 3: R P Q Q S L V Q N Y P H Y A L L Q D A I

env

```

1:
2:
3: N S M I G G C T T U G Y E I N S F F

1981      2001      2021
AATAAAGAGGTGAAACAGATTAATTCAGACATCCAGGTACTGAGCTAAGAT
1: N K R G E T P H C Q T S Q U Y W N Q G
2: Y K E U K Q T I U K H P R Y T G T N N
3: Q P K F C N F F I S N I P G I L E L

2041      2061      2081
TATTAAGATCAATTAAAGGCTTCAGAGGAGAGATCCGGAAGTTACGTTGATGAG
1: Q Q N Q F N G F E E R R S G S Y L H U U
2: E K I N L T A P R G G D P E U T F M W Y
3: C I K S I Q R L L E E E I F K C P E C C

2101      2121      2141
AAGTTGAGAGGAGAGGATTCCTCTGCTGTAAGATGAGTTGGTTTCTAAGTTGGGTAGAG
1: K I Q R R U P L L Q N E L U S M L G P
2: N C R G E F L Y D K M N W E L N W U E P
3: R I A E E S S E T U K F I S F C I G P F

2161      2181      2201
TCTGATGTAAGTACCCAGAGGCGAAGGACGCGATAGAAAGCAATTACGTCGCGTGA
1: G G C N Y P E A K G T A Q K E L P A U Q
2: P D U I T Q E P K E R H R R N Y U P C H
3: I E M Q L P R E F R N G I E G I T C S D

2221      2241      2261
TATTCAGAGATGAATCAACATTCGATAGAGTAGGCAAGATGTTTATTTGCTTCAG
1: Y Q T N N Q H L A Q S R Q K C L F A S
2: I E Q I I N T W H K U G K N U Y L F F
3: J L D K Q S T L S I K Q A K R T J C I

2281      2301      2321
AAGAGGAGACCTTATGTTGATAGCTTCAGTACCACTGTCATAGCAAGATAGATTGAG
1: P G R P H L Q L H S D Q S H S H R L U
2: E G D L T C N T U T S I I A N I R U T
3: E S E T S R U T U Q Q P U S E Q T Q I F

2341      2361      2381
TGATGTAAAGCAAGATAGTATCAAGTGGATGAGAGCTGGCAGAACCTATGATTC
1: P W K P N Q Y H H E C R G G R T U S I G
2: C G N Q T S I T M S A E U A E I Y P I E
3: L M E T K L U S P Q U Q P W Q N C I D W

2401      2421      2441
ATTAGGAGATTAGASATTACTATAGATGATTCGATTGCTTGGCTGCACAGATAT
1: A Q R L Q T Q P P H S T W L G P H F C
2: L G R Y K L U E I T P I G L A P T Q U
3: L L I Y N E T S L F L A W F P Q P

2461      2481      2501
GAGGTACACTACTGCTGGCACTGAGCAAGATAAGAGAGGGGCTTTTGTCTAGCTT
1: E L H Y W W L N K Q Y P E L C A R U
2: G T L S T G T G T G T G T G T G T G T
3: E T L L U Y W G Y K E S S I S G E

2521      2541      2561
GATTTGCTGCAAGCTGAGTTCTGCAATGGGCGGAGCTGCTGACGCTGACCA
1: G I S F P G R F C N G R D U U D A D R
2: G F L A T A C S A H G A T S L T L T A Q
3: W U E S Q P C U L G W G R R R Q R Q P Y

```

env

3121 3201 3281  
TGGAGAAAGGCGGTGGCAACAGCTCCTTGGCGCTTGGCAGATAGAAATATATTTCATTTCCTGAT  
1: W R R R W Q Q L L A L A D P I Y S F P D  
2: G E G G G N S S W P W R I E Y I H E L  
3: V E K A V A T A P G L G R I C N I E I S P

env

First premature stop  
codon env

rev  
env  
tat

$$\frac{25}{25} - \frac{25}{25}$$
$$\frac{12V}{ENV} = \frac{12V}{ENV}$$

3301 3321 3341  
 AGGATACAGATCTTCCAACTACTGAGAGAGGCTCTTGGACCCCTGATAT  
 1: S I P S P P T N T P E A L C D P T N  
 2: A Q T L Q P I L G P A T L K  
 3: E H T E S S N Q Y S P G S L R P Y E C F

3361 3381 3401  
 AGGATACAGATCTTCCAACTACTGAGAGAGGCTCTTGGACCCCTGATAT  
 1: R S P G T D L P T I V E L E P S Q  
 2: E U L E T E L T Y L Q Y R W S Y F H I T  
 3: E K S S G L N Q P Y N I G G A I S M M

3421 3441 3461  
 AGGATACAGATCTTCCAACTACTGAGAGAGGCTCTTGGACCCCTGATAT  
 1: G P S R L E I C D R N S C G R V G S M  
 2: V G A G W R S A T E T L A G A W G I L  
 3: P S K P A G D L R Q K L L R A P E F T T

3481 3501 3521  
 AGGATACAGATCTTCCAACTACTGAGAGAGGCTCTTGGACCCCTGATAT  
 1: S P S E R W K I D P R N P P E S S M  
 2: E T L F R G G R I L A P R R I L Q M  
 3: E R I L G E V E D P S S S S L G G I D M

3541 3561 3581  
 AGGATACAGATCTTCCAACTACTGAGAGAGGCTCTTGGACCCCTGATAT  
 1: A P A H A L V P D P N T P S S I R  
 2: L K L T L Q G T E I C S G A V Y C M C  
 3: S L S S E S C E G D K Y N G S G S M

3601 3621 3641  
 AGGATACAGATCTTCCAACTACTGAGAGAGGCTCTTGGACCCCTGATAT  
 1: H G C T Q L K K R K A Q T E N M I L M  
 2: M E V P S G P R K S K J S T G K T Y Y M  
 3: P A R A P A E E K E L A Y P P P T

3661 3681 3701  
 AGGATACAGATCTTCCAACTACTGAGAGAGGCTCTTGGACCCCTGATAT  
 1: I P M R K M M T W Q G Y Q G S G R P P  
 2: Y R E S R Q L G R C I S E A K S S S M  
 3: I I D E E D N D L V G V S V P R K P T

3721 3741 3761  
 AGGATACAGATCTTCCAACTACTGAGAGAGGCTCTTGGACCCCTGATAT  
 1: E C P T N W Q Q T C L I L I G G  
 2: S D L Q I G N S Y V S F Y R P C L M  
 3: T A M T Y K L A T D M S H E I K T G

3781 3801 3821  
 AGGATACAGATCTTCCAACTACTGAGAGAGGCTCTTGGACCCCTGATAT  
 1: W K G P I T V Q E D T E S S T C T P K W  
 2: C T C T C T C T C T C T C T C T C T C T  
 3: L E G I L S S R R R R R R R R R R

3841 3861 3881  
 AGGATACAGATCTTCCAACTACTGAGAGAGGCTCTTGGACCCCTGATAT  
 1: K K A S Y D I G R I T P D I Q L I T  
 2: P P H H T R L A G L H L R T R R I L M  
 3: E E G I I P D W Q D Y T S E P C I P M



